cited in the European Search Report of EP 04 00 6542.7
Your Ref.: 8031-1033

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 665 693 B1

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent:
15.03.2000 Bulletin 2000/11

(51) Int. Cl.[7]: **H04N 7/50**, H04N 7/56

(21) Application number: **94120862.1**

(22) Date of filing: **28.12.1994**

(54) Dynamic bit rate control system for very low bit rate video and associated audio coding

Dynamisches Bitratensteuerungssystem für Video und entsprechende Audiokodierung sehr niedriger Datenrate.

Système de contrôle dynamique de débit de bits pour un codage à très bas débit de vidéo et audio associée

(84) Designated Contracting States:
DE FR GB IT NL

(30) Priority: **28.12.1993 JP 33697593**

(43) Date of publication of application:
02.08.1995 Bulletin 1995/31

(73) Proprietor:
MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.
Kadoma-shi, Osaka-fu, 571 (JP)

(72) Inventors:
- **Huang, Si Jun,**
  BLK 525,
  13-404, Singapore 1646 (SG)
- **Tan, Ah Peng,**
  BLK 542,
  10-03 Singapore 1955 (SG)

(74) Representative:
Grünecker, Kinkeldey,
Stockmair & Schwanhäusser
Anwaltssozietät
Maximilianstrasse 58
80538 München (DE)

(56) References cited:
EP-A- 0 501 699
EP-A- 0 515 101
EP-A- 0 540 961
JP-A- 4 360 481

- **SIGNAL PROCESSING IMAGE COMMUNICATION., vol. 4, 1992 AMSTERDAM NL, pages 153-159, XP 000273161 MACINNIS 'The MPEG systems coding specification'**

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 0 665 693 B1

Printed by Xerox (UK) Business Services
2.16.7/3.6

BNSDOCID: <EP_____0665693B1_I_>

Description

[0001] The present invention relates to very low bit rate video and associated audio coding, specifically to the dynamic allocation of video bit rate according to the instant bit rate consumption of audio signal in a constant bit rate system.

[0002] Digital video and associated audio coding plays a key role in the industrial applications of digital signal storage, processing, transmission and distribution. Various digital coding standards have been and are being developed by various standardization bodies. For the moving picture and associated audio coding, typical coding scheme involve three parts, namely, video coding, audio coding and system multiplexing.

[0003] The current transform coding algorithms adopted by the MPEG phase one and phase two standards involve such techniques as transformation, quantization and variable length coding. To increase the coding efficiency, predictive coding techniques such as inter picture prediction,
motion estimation and compensation are used. Therefore, a picture could be encoded by reducing its spatial redundancy within the picture itself. A picture coded in this way is usually referred to as an I picture. A picture could also be encoded in inter picture mode. If a picture is predicted only from the previous picture, it is referred to as a P picture. If a picture is predicted both from previous picture and future picture, the picture is referred to bidirectional predictive coded picture, which is usually called a B picture. A main feature of the current video coding scheme is that the bit rate of the coded bitstream is a fixed constant by using a FIFO buffer. The fullness of the buffer is used to adjust the quantization step and bits available for the I, B and P picture so as to control the bit rate of each coded picture.

[0004] Many audio coding algorithms have emerged with the introduction of digital communication links, mobile communication, entertainment and multimedia services. In particular, various very low bit rate speech coding algorithms have been standardized to reduce transmission bit rate or the memory capacity of the voice storage systems. Toll quality speech could be obtained from the recently standardized CCITT G.728 16 kbits/s codec. Communications quality could be obtained using the USA Federal Standard 1016 4.8 kbits/s voice coder. The Vector Sum Excited Linear Predictive Coder standardized for North American and Japan cellular communications could provide near toll quality at 8.0 kbits/s operation. These algorithms generally fall into the class of speech coders known as Code Excited or Vector Excited Linear Prediction coders and have been typically designed for constant bit rate transmissions. The number of bits per audio frame is also kept constant.

[0005] Very limited variable bit rate for speech coding has been considered for packet switching networks, digital speech interpolation systems and digital communication multiplication equipment systems.

[0006] The main function of the system encoder is to provide the necessary and sufficient information to synchronize the decoding and presentation of video and audio information, and at the same time ensuring that coded data in the decoders' buffer do not overflow or underflow. Coding the system layer information includes packetizing the data into packets and creating time-stamps for the packet header. Two time stamps are used: presentation time stamp (PTS) which indicates when the presentation unit of an audio frame or video picture should be play or broadcast, and decoding time stamp (DTS) which indicates the time to decode an audio or video picture. The PTS and DTS have a common time base, called system clock reference (SCR), to unify the measurement of the timing, ensure correct synchronization and buffer management. In a fixed constant bit rate environment, the system expects a fixed bit rate audio and video for the correct buffer management.

[0007] The coding scheme described above has following problems that prevent us from achieving a very high compression ratio and coding efficiency, which are the key factors for various very low bit rate video and associated audio coding applications.

[0008] The inefficiency of current coding scheme arises from:

1) Constant audio bit rate

[0009] The current audio coding scheme encodes and audio signal in constant bit rate. If we define the audio information as a complexity measure which reflects the signal intensity and frequency distribution, we can find the complexity various with time. For example, considering the videophone application, when one party of the communication is talking, the other party is usually listening and silent. This mean that there are moments when the listener does not make any voice input. Secondly, there are silent moments existing even in person's talking. A constant audio encoder wastes the bandwidth at those silent moments. For very low bit rate coding applications, it is possible that the audio occupies similar or even higher bandwidth than the video. The problem to be solved here is to use a variable bit rate audio encoder for audio coding and save the bits for video coding use.

2) Constant video bit rate

[0010] The current video coding scheme provides a constant bitstream output by using a FIFO buffer at the end of the encoder. The instant bit usage and buffer fullness are used to adjust the bits for each picture and the quantization

BNSDOCID: <EP_____0665693B1_I_>

step. The latter is used to control the bit rate of the next encoded macroblock within the picture. This bit rate control process is done within the video encoder itself, which is independent of the bit usage of the audio encoder. How to make use of the bit saving from the audio encoder to improve the video coding quality is the problem to be solved.

3) Video dynamic bit rate control

[0011] The current video coding scheme implements the bit rate control by allocating a certain number of bits to each picture and also adjusting the quantization step for each macroblock of a picture. There are cases when there are sufficient bits available that are more than enough to adjust the quantization step to its minimal value. Under such situation, the extra bits could not be efficiently used.

[0012] Reference US-A-5,309,290, (published on 03.05.1994), corresponding to JP-A-4360481, describes a digital video-audio recording and reproducing apparatus wherein both, the view signal encoder and the audio signal encoder having a variable reduction rate. The reduction rate in both of the encoders are controlled in such a manner that the sum of the information amount of the video signal and the audio signal is maintained constant. A reference amount provided for the encoding procedures of the video and audio signal respectively are generated by a controller. The provided reference amount is based on the difference between values fed back from the audio encoder and video encoder which represent a fair amount of information in the respective audio or video recording, and the reference amount of information for the video or audio signal. The amount of information used by the respective encoder is provided by information amount estimation circuits.

[0013] Another problem is whether the minimal quantization step is the best way to encode a picture in the predetermined encoding mode, i.e., I picture, P picture and B picture. Especially when the current picture is set to be encoded in B picture, will the smaller quantization step give better coded picture quality than a P picture or even an I picture if the amount of bits available allows us to do either a P picture or an I picture coding? The above discussion could be concluded as a problem of how to dynamically select I, P and B picture coding modes.

[0014] A second problem under this item relates to very low bit rate coding where the picture rate of the coding scheme is usually less than the rate required by real time video. For the videophone application, as an example, the picture rate is usually set around 10 pictures per second. Because of this reduced picture rate, the jig effects, would appear if the objects in the picture moves too fast. This problem also makes the predictive coding difficult when a scene is changed. How to efficiently make use of the available bits to insert a picture to the pre-determined picture rate is another issue to be addressed to improve the picture coding quality.

4) System multiplexer

[0015] Current system multiplexer accepts only the constant bit rate audio and video bitstreams to multiplex them into a constant bit rate bit system bitstream. There is no mechanism to control the audio and video encoder for the dynamic bandwidth allocation between the two encoders.

[0016] It is the object of the invention to provide a coding system for video and audio signals enabling a more efficient video encoding.

[0017] This is achieved by the features of claim 1.

[0018] In particular it would be advantageous to provide a communication link between the audio and video encoder such that the bit saving obtained in audio coding could be efficiently used by the video encoder to improve picture quality. It is also important to provide more flexible video bit rate control mechanism for variable I, B, P picture coding modes and dynamic picture insertion.

[0019] A dynamic bit rate control video and associated audio coding system has been invented. The coding system at first codes a few frames of audio. This number of audio frames is decided by the picture rate of video and audio frame size. It then counts the bit spending and calculates the bit saving from the audio encoder during the previous picture coding time period. By monitoring the input audio level and analyzing of its frequency and energy contents, maximum bits may be saved during silence period. For non-silence period, the number of bits to be used for coding of each audio frame is selected such as to attain a predetermined frequency-weighted Signal-to-Noise measure or other perceptual measures which assures the desired audio quality level.

[0020] The saved bits from the audio encoder is sent to a bit rate allocator, which makes use of these bits together with the bit consumption of the coded previous picture and the fullness of a first-in-first-out (FIFO) buffer to calculate the available bits for the current picture coding.

[0021] The available bit amount for the current picture is sent to a coding mode selector in the video encoder. This coding mode selector decides the I, B and P picture coding for the current picture. It also decides whether an extra picture is to be inserted when the picture rate is lower than the normal picture rate required by real time video application. The available bits are also used to control the quantization step in a similar way to the current coding scheme when the picture coding mode is selected.

BNSDOCID: <EP_____0665693B1_I_>

[0022] The system multiplexer will make up for the video delay with respect to the audio coding by inserting in the system packet the correct time stamp for the respective video and audio data. The synchronized audio and video bitstream will then be multiplexed to form a constant bit rate system bitstream.

[0023] The operation of the coding system with dynamic bit rate control of combined audio and video coding are illustrated in Figure 1. the coding system involves an audio encoder 2, a video encoder 10, a first-in-first-out (FIFO) buffer 11, a bit rate allocator 5 and a system multiplexer 14. This invention covers not only such a system configuration and operation, but also covers new means needed by the audio encoder 2, video encoder 10, bit rate allocator 5 and the system multiplexer 14 to support the operation of the overall coding system.

[0024] In Figure 1, the audio bitstream 3 is encoded by a variable bit rate audio encoder 2, which includes an audio detection and identification (ADI) module 103 (Figure 5) that performs several tasks, namely,

a) to detect the silence moments in the audio signal and makes decisions for skipped audio frames
b) to classify non-silence audio segments into predominantly voiced or unvoiced segments
c) to allocate bits based on the frequency-weighted energy levels of the analyzed speech segments
d) to transfer bit spending information to a bit saving counter (BSC) 25 (Figure 4).

[0025] The BSC 25 calculates the bit saving after each audio frame is coded. Once the BSC 25 receives a request signal from the bit rate allocator (BRA) 5, it calculates the bits saved by the audio frames during the previous picture encoding period. This bit saving amount 4, which is denoted as B_s, tells the BRA 5 the amount of bits contributed by the audio encoder 2 for video encoding use.

[0026] At the same time, the bit amount used by the previous picture, B_u 7, is provided by the video encoder 10 to the BRA 5. Similarly the fullness of the FIFO buffer 11, B_f 9, is also provided by a fullness detector in FIFO buffer 11. These three parameters, B_s, B_u and B_f, are then used by the BRA 5 for the calculation of the bits available, B_a 8, for the current picture coding.

[0027] According to the value of B_a 8 provided by the BRA 5, the video encoder 10 selects the various picture coding modes to achieve the best possible picture coding quality. The output 12 of the coded picture will be sent to the FIFO buffer 11. The bit usage 7 of the current picture is reported to the BRA 5 for the use of the next picture coding.

[0028] The coded audio 3 and video 12 bitstreams are then multiplexed by a system multiplexer 14 where the audio and video synchronization is realized. The system multiplexer 14 multiplexes the audio and video in a constant bit rate and produces a constant system bitstream 15 to either a communication channel or a digital storage media.

Figure 1 shows the block diagram of the dynamic bit rate controlled video and associated audio coding system.

Figure 2 illustrates the time and bit rate relationship during audio and video encoding process.

Figure 3 illustrates the time and bit rate relationship during the multiplexing process.

Figure 4 shows the relationship between the bit saving counter and the variable bit rate audio encoder.

Figure 5 shows a block diagram of a variable bit rate audio encoder.

Figure 6 shows a block diagram of a configuration of a variable bit rate video encoder together with a coding mode selector and a bit consumption counter.

Figure 7 illustrates the mechanism of dynamic picture coding mode selection flow chart.

Figure 8 shows a block diagram of a video transform coding scheme.

Figure 9 illustrates the FIFO buffer fullness.

1. Bit rate dynamic control mechanism

[0029] A preferred embodiment of the coding system is shown in Figure 1 where the coding system involves only two picture coding modes. The first picture coding mode is the intrapicture coding that generates I pictures. The second picture coding mode is the predictive coding that generates p pictures. the prediction is only from the previous picture. However, the coding system of this invention is not only limited to I and P picture coding modes. It can also be applied to the I, P and B picture coding modes:

[0030] The control mechanism of the dynamic bit rate allocation is shown in Figure 2. Firstly, there is a preset maxi-

BNSDOCID: <EP_____0665693B1_I_>

mum bit rate for audio coding, which is denoted as B_{amax}. The audio signal 1 is encoded by using a variable bit rate encoder 2 with the constraint that the coded audio bit rate shall not exceed the B_{amax}. Assuming the audio frame rate is set at F_a frames per second and the video picture rate is set at F_v pictures per second. The audio encoder 2 shall start encoding the audio signal 1 for a time period T_d. This time is the delay needed by the video encoder 10 and it is set to one picture period. Immediately before starting the encoding of the first picture, the BRA 5 shall send the B_a 8 to the video encoder 10 to specify the available bits for the current picture coding. The video encoder 10 analyzes the value of B_a 8 so as to apply proper picture coding modes.

[0031] The effect of applying the above rate control scheme is also shown in Figure 2. If conventional coding scheme is used, which provides a constant bit rate video bitstream output, the average bit rate would be the amount denoted as B_{vavg}. The first I picture spends much more bits than the succeeding P pictures (It is usual that the I picture bit rate is twice as high as the P picture bit rate). With the invented coding scheme, the bit saving B_s 4 of the audio frames during the previous picture period is used by the current picture. This results in that each picture is able to be encoded with a certain amount of extra bits B_s 4 that are saved by the audio encoder 2. Sometimes, if there is a silent period in audio signal, the whole bandwidth reserved for audio can then be used by video, as the case of the 8th picture in Figure 2. These added bits actually helps video encoder to improve the coded picture quality. It is obvious to see from the Figure 2 that the average video bit rate B'_{vavg} is increased compared to the original video bit rate average B_{vavg}. Following formula gives the relation of the average video bit rate and audio bit rate.

$$B'_{vavg} - B_{vavg} = B_{amax} - B_{aavg}$$

[0032] Where the B_{aavg} is the average audio bit rate.

[0033] Because the video encoder 10 starts the encoding process with a delay T_d with respect to the starting time of audio coding and also because the video bit rate is dynamically allocated according the audio bit saving during the time period of the previous picture, the coded audio bitstream 3 shall be delayed with a same time T_d to match the corresponding video bitstream 12 in both time and bit rate. This operation is illustrated in Figure 3, where the bit saving in audio and bit spending in video are matched in time. This alignment control is achieved by the system multiplexer 14 to selectively take audio and video bitstreams from the FIFO buffer 11 and put in the correct time stamps for audio and video synchronization.

[0034] From both Figure 2 and Figure 3, a distinguish feature that is different from the conventional audio and video coding system is that both the audio bitstream 3 and video bitstream 12 are of variable bit rates. However, when the audio bitstream 3 and video bitstream 12 are synchronized by the system multiplexer 14, the output system bitstream 15 is in constant bit rate.

2. Variable bit rate audio encoder

[0035] The relationship between the bit saving counter and the variable bit rate audio encoder 2 is depicted in Figure 4. A bit saving counter (BSC) 25 is introduced. As it is defined above, the audio signal 20 is segmented into F_a frames per second. Assuming the bit spending of the ith audio frame is $f_a(i)$ bits 23, the bit saving during one picture period could be calculated by the following formula:

$$B_s = B_{amax} \cdot N_{af} - \sum_{i=1}^{Naf} a\,(i)$$

[0036] Where the N_{af} 28 is the number of audio frames that have been encoded before the B_s request was made from the BRA 5. Once the B_s 29 is sent to the BRA 5, the audio frame counter 27 is reset to zero and continue counting for the next picture period.

[0037] As illustrated in Figure 5, for the purpose of determining the bit spending for each audio frame, the input speech signal 101 is passed to the prediction analysis module 102 and audio detection and identification module 103. The prediction analysis module 102 performs the short-term linear spectrum analysis using a specified Mth order autocorrelation analysis. In the audio detection and identification module 103, the energy distribution across the audio frequency band is computed. Where silence is detected based on a threshold measure, no coding of audio frame need be done. Together with the resultant prediction coefficients 104 from prediction analysis 102, the characterization of the speech segment into voiced or unvoiced components (or more subclasses) would take place. Depending on the type of applications, a maximum bit rate usable for coding of each audio frame could first be present in the bit rate determination module 106. From the prediction coefficient and energy distribution information, a discrete bit rate level equal to or less than the maximum bit rate usable would be selected in the bit rate determination module 106, based on a fre-

BNSDOCID: <EP_____0665693B1_I_>

quency-weighted Signal-to-Noise measure or other perceptual measure for coding of the audio frame. The bit spending data 116 from each audio frame will be transmitted to the bit saving counter. The audio frame number 115 generated from a cyclic counter would be transmitted to the frame counter. Coding at a lower bit rate than the maximum could be achieved by using proper subsets of the codebooks 108 (or by reduced codebooks or by re-design for the parametric representation). The multiple codebooks 108 constitute the excitation generators which can generate a variety of sequences including pulse-like and noise-like sequences. The summation of the excitation sequences 109 is passed through a predictor which can consist various predictors 110 e.g. pitch predictor and spectral predictor. The predicted signal 111 is subtracted from the input speech 101 resulting in a difference or error signal 112 which is then perceptually weighted. The perceptually weighted error 114 is then used to drive the error minimization Vector Quantization search procedure. The line spectral pairs 118 and appropriate parametric codes 117 would be transmitted as components of the audio bitstream.

[0038] The variable bit rate operation considered thus far has a different approach from the algorithms which have a nominal rate but can operate at bit rates higher and lower than the nominal rate.

3. Bit rate allocator (BRA)

[0039] As illustrated in Figure 1, the BRA 5 receives the audio bit saving B_s 4 from the audio encoder 2, the video bits used by the previous coded picture B_u 7 from the video encoder 10 and the fullness of FIFO buffer B_f 9. Assuming that the preset target bits for picture j is $B_{t(j)}$, which is equal to the preset I picture coding bit rate B_{tI} or the preset P picture coding bit rate B_{tP} for I picture coding and P picture coding respectively, the available bits for the j th picture coding shall be calculated by the following control means.

$$B_{a(j)} = B_{t(j-l)} - B_{u(j-l)} + B_{s(j-l)} + B_{t(j)}$$

$$if(B_{a(j)} + B_{f(j)} + B_{amax} > B_{fmax}) \{ B_{a(j)} < B_{fmax} - B_{f(j)} - B_{amax} \}$$

$$if (B_{a(j)} + B_{f(j)} + B_{amax} < B_{Av}) \{ B_{a(j)} > B_{Av} - B_{f(j)} - B_{amax} \}$$

[0040] Where the B_{fmax} is the FIFO buffer size and B_{AV} is the constant system bitstream bit rate. The available bits for the jth picture coding is equal to the difference of the target bits for and the actual bits used by the (j-l)th picture, plus the bit saving of audio encoder 2 during the (j-l) picture coding period, plus further the target bits for the jth picture according to the default coding mode. If the sum of the available bits $B_{a(j)}$, the current FIFO buffer fullness $B_{f(j)}$ and the maximal audio bit rate B_{amax} is greater than the maximal FIFO buffer fullness B_{fmax}, the available bits for the picture j shall be calculated as the maximum FIFO buffer fullness B_{fmax} minus the current buffer fullness $B_{f(j)}$ and minus the maximal audio bit rate B_{amax} to avoid the FIFO buffer 11 overflow. If the sum of the available bits $B_{a(j)}$, the current FIFO buffer fullness $B_{f(j)}$ and the maximal audio bit rate B_{fmax} is less than the combined audio and video bit rate B_{AV}, the available bits for the picture j shall be calculated as the B_{AV} minus the current buffer fullness $B_{f(j)}$ and minus the maximal audio bit rate B_{amax} to avoid the FIFO buffer 11 underflow.

4. Variable bit rate video encoder

[0041] The variable bit rate video encoder 10 is configured together with a coding mode selector 31 and a bit consumption counter 35 as illustrated in Figure 6. The available bits for the current picture B_a 30 is sent to a coding mode selector (CMS) 31. Depending on the value of B_a 30, the CMS 31 conducts a decision making process to select I, P coding or an extra P picture coding. This decision is provided for a variable bitrate video encoder 33 to control the coding of the current picture 37. The encoded video bitstream is output through link 34. The output 34 is also sent to the bits consumption counter 35 so that the number of bits used for the current picture coding B_u 36 is calculated and sent to the BRA 5 (Figurel) for the next picture coding use.

[0042] The decision making process of CMS 31 is shown in flow chart in Figure 7. Assume that the minimal bits needed to encode an I picture is B_{Imin} and the minimal bits needed to encode a P picture is B_{Pmin}. Assume further that there is a present threshold TH_I to check the picture difference (PD) between the previous encoded picture and the picture that is located in the middle of the previous encoded picture and the current picture to be encoded.

[0043] The available bits for the current picture B_a50 is input to the comparator 51. If the B_a is greater than two times B_{Pmin}, it implies that the available bits are enough to encode two P pictures. In other words, an extra P picture could be inserted. In this case, the output of the comparator 51 goes through the link 53 to the comparator 54, which is used to check further whether it is necessary to insert a P picture based on the PD value as defined previously. If the PD value is greater than the preset threshold TH_I, it implies that there is substantial changes of picture contents from the previous encoded picture to the picture in the middle of the previous encoded picture and the current picture to be encoded.

BNSDOCID: <EP_____0665693B1_I_>

Therefore, the decision 62 is made to encode two P pictures if the output of the comparator 54 is "Yes" through the link 58.

[0044] However, if the comparator 54 outputs a "No" through link 59, it implies that the middle picture has no substantial changes with respect to the previous encoded picture and there is no need to insert an extra P picture. The available bit amount B_a is then check through link 59 by the comparator 55 for the decision making of either an I picture coding or a P picture coding. This process also applies to the case when the comparator 51 outputs "No" from the link 52. Therefore, the succeeding decision making procedure applies to both cases of link 52 and link 59.

[0045] In the comparator 55, the available bit amount B_a is compared with the minimal I picture coding bit amount, B_{Imin}.

[0046] if $B_a \geq B_{Imin}$, it implies that the available bits are enough to encode an I picture and the decision 60, I picture coding, is activated by the output from the link 56. If $B_a < B_{Imin}$, the decision 61, P picture coding, is selected by the output from the link 57.

[0046] Only one of the decisions of 60, 61 and 62 is selected at one time and sent through one of the links of 63, 64 and 65 to the coding mode encoder 66, which converts the selected decision into an unique coding mode code to the output link 67.

[0047] The mechanism described above is defined as the dynamic I and P picture coding mode control.

[0048] When a picture coding mode is determined, the B_a is used to adjust the quantization step so as to dynamically control the bit rate generated from the current picture coding. This mechanism is defined as dynamic quantization step control.

[0049] A variable bit rate video encoder 33 is then used to code the current picture according to the picture coding mode 32 selected by the CMS 31. One preferred embodiment of the video encoder 33 is shown in Figure 8, which is a transform coding system.

[0050] An input video signal contains pixel data of each picture is input and stored in an input picture memory 71. A block sampling circuit 72 receives the data stored in the input picture memory 71 through a line 84 and partitions the picture data into spatially non-overlapping blocks of pixel data. To provide a reasonable level of adaptiveness, a block size of 8x8 pixels may be used. The picture number of the picture is also stored in the input picture memory 71 and passed through the block sampling circuit 72 to a switch (SI) 73. Based on the picture number, the switch 73 selectively delivers the output blocks of pixel data from the block sampling circuit 72 through a line 86 for coding in the intra-coded mode or to a line 87 for coding in the predictive-coded mode.

[0051] For the intra-coded mode, the output of the block sampling circuit 72 is passed through line 86 to a discrete cosine transform (DCT) circuit 76. The DCT circuit 76 performs discrete cosine transform, which is a popular mathematical transformation for converting image data to data in the frequency domain. The transformed data, i.e. DCT coefficients, are then subjected to the process of quantization in a quantization circuit 77 using a quantizer matrix and a quantization step size which is given by a rate controller 79 through line 93. The quantized data are passed together with the quantizer step as side information, through line 91 to run-length coding & variable length coding circuit 78 which performs run-length coding of the quantized data and the associated side information followed by variable length coding of the run-length coded result. The output of the run-length coding & variable length coding circuit 78 is a coded bit stream 100 ready to be transmitted to the decoder. This bit stream is also passed through line 92 to rate controller 79. Based on the number of bits already used at the time of encoding the block, the rate controller 79 adjusts the quantizer step so that the output bit stream satisfies the bit rate requirement of the encoder system. The quantized values obtained by the quantization circuit 77 are passed also through a line 94 to an inverse quantization circuit 80 and an inverse discrete cosine transform (inverse DCT) circuit 81. The inverse quantization circuit 80 and inverse DCT perform the reverse process performed by the DCT circuit 76 and quantization 77 to obtain reconstructed data. The reconstructed data is stored in a local decoded picture memory 82 through line 96, and will be used for motion estimation and compensation process for the next input picture, which will be a predictive-coded picture. After the whole of the I-picture has been coded, the rate controller will calculate the total bit used for coding the I-picture, and decide the number of bit to be allocated to the next P and B-picture.

[0052] For the predictive-coded mode, the output of the block sampling 72 is passed through switch S1 and then a line 87 to a motion vector (MV) estimation circuit 74. In MV estimation circuit 74, motion vector estimation is performed to determined the MV from the adjacent picture using the local by decoded picture from the local decoder picture memory 82, obtained through line 98. The MV estimation circuit 74 finds the closest match by, for example, determining the direction of translatory motion of the blocks of pixels from one picture to the next by finding the best matching block based on some predetermined criteria, such as mean square error.

[0053] The MV obtained in the motion vector estimation process are passed, together with the blocks to be code, through line 88 to a motion compensation circuit 75 which performs motion compensation using the blocks stored in the local decoded picture memory 82 as predicted blocks. The predicted block is obtained from the local decoded memory using the offset given by the MV. Differences between values of the pixels of the block to be coded and values of the pixels of the predicted block are computed to obtain a differential block composed of the difference values. Based on

BNSDOCID: <EP_____0665693B1_I_>

the differential block, the motion compensation circuit 75 also determines whether the block to be passed to DCT circuit 76 for transformation (block to be transformed) should be motion compensated or not by comparing the activity of the block to coded and the differential block using some predefined criteria, such as standard deviation of the blocks. If the block to be transformed is decided to be motion compensated, its differential block is passed through line 89 to DCT circuit 76. If the block to be transformed is decided not to be motion compensated, the block itself (i.e. the original block outputted from the block sampling circuit 72) is passed through the line 89 to the DCT circuit 76 (which is the same as the block of the intra-coded picture). Further, the motion compensation circuit 75 produces an information bit indicating whether a block is coded with motion compensation or not and passes the information bit together with the motion vector to the DCT circuit 76. That is, the data outputted from the motion compensation circuit 75 through the line 89 to the DCT circuit 76 are either the differential block (coded with motion compensation) or the original block (coded without motion compensation), the information bit indicating whether or not the block is coded with motion compensation, and also the motion vector in the case of differential block.

[0054] The block data from the motion compensation circuit 75 are transformed to DCT coefficients in the DCT circuit 76, and the DCT coefficients are quantized in the quantization circuit 77 to be quantized DCT coefficients. The information bits each indicates whether or not a block is coded with motion compensation and the MV are passed through the DCT circuit 76 and the quantization circuit 77 through a line 90. The quantized data (quantized DCT coefficients), together with side information including the quantization step, the motion vectors and the information bits each indicating whether or not motion compensation is done to a block are passed through the line 91 to the run-length coding & variable length coding circuit 78 to be subjected to run-length coding and variable length coding to decode the output coded bit stream. The bit stream is also passed through the line 92 to the rate controller 79. The quantized data from the quantization circuit 77 is also passed through the line 94 to be subjected to inverse quantization in the inverse quantization circuit 80, inverse DCT in the inverse DCT circuit 81 and stored in the local decoded picture memory 82 for the encoding of the next picture.

5. FIFO buffer

[0055] Figure 9 shows an example of the FIFO buffer status during encoding process of audio and video. For the first time period T_d, the coded audio bitstream 3 is written to the FIFO buffer 11. Starting from the moment T_d, the FIFO buffer 11 is filled by both audio bitstream 3 and the video bitstream 12. When the FIFO buffer 11 is filled up to a preset fullness B_O, the system multiplexer 14 starts to collect bits from the FIFO buffer 11. Since the system multiplexer 14 takes bits from the FIFO buffer 11 in a constant bit rate B_{AV}, the buffer fullness is reduced at a fixed amount, denoted as B_{out}. This value remains for all the time. Because the audio and video bit rate are variable, the combination of them is also a variable bit rate. Therefore, the speed of FIFO buffer 11 is changing from one time period to another, as illustrated in Figure 9 by different gradient of the buffer fullness curve in each time period. The B_O is set to avoid the FIFO buffer 11 overflow or underflow. However, this also introduces a time delay for the bitstreams to be transmitted, as illustrated in Figure 9 by $3T_d$.

6. System Multiplexer

[0056] The system multiplexer multiplexes the input video and audio data from the FIFO buffer by packetizing the input video and audio data into packets of fixed length, and inserting the time stamp into the corresponding audio frames and video picture to ensure the correct synchronization. Two time stamps are used: presentation time stamp (PTS) which indicates when the presentation unit of an audio or video picture should be played or broadcast, and decoding time stamp (DTS) which indicates the time to decode an audio or video picture. The PTS and DTS have a common time base, called system clock reference (SCR), to unify the measurement of the timing, ensure correct synchronization and buffer management.

[0057] For audio data, the system multiplexer keep track of the number of pictures sent and as new data are obtained from the FIFO, an appropriate PTS will be inserted to the packet header. No DTS is necessary as the presentation time has a fixed relationship with the decoding time.

[0058] For video data, the system multiplexer also keeps track of the number of pictures sent and appropriate PTS and DTS are inserted to the packet header. DTS are necessary in the case when the decoding time and the presentation time of the picture are different.

7. Effects of the Invention

[0059] The present invention has significant improvement over the prior art method.

[0060] The invention is a constant bit rate video and associated audio coding system which has various dynamic control mechanisms that would allow for saving in the audio encoder to be used by the video encoder.

BNSDOCID: <EP_____0665693B1_I_>

[0061] The invention provides the mechanisms to control the picture rate as well as picture coding modes. These result in more efficient use of bits available over a fixed period of time, thereby improving the quality of video signal significantly without compromising on the audio quality.

[0062] The invention is especially effective for very low bit rate video and associated audio coding. For an instance, when this invention is applied to the videophone application, it can improve the picture quality by making use all of the audio bit rate of the listener so that the pictures are encoded in more bits and transmitted to the person who is speaking.

Claims

1. A coding system for encoding a video signal (6) and an audio signal (1), comprising:

   a variable bit rate audio encoder (2) for encoding said audio signal (1) and providing an amount (4) of saved bits,
   a variable bit rate video encoder (10) for encoding said video signal (6),
   a FIFO buffer (11) for buffering the encoded audio bit stream (3) and video bit stream (12) and providing a constant rate bit stream (13), and
   a bit rate allocator (5) for dynamically allocating an amount (8) of bits for a current picture of said video signal (6) to be encoded, said bit rate allocator (5) calculating said amount (8) of bits for the current picture depending on the amount (4) of said saved bits of the audio encoder (2), on the amount (7) of bits used for a previous picture of said video signal (6) and on a fullness (9) of said FIFO buffer (11),
   a multiplexer (14) for multiplexing the constant rate audio and video bit stream (13) with added time stamps for a synchronisation of the audio signal and the video signal.

2. A coding system according to claim 1 wherein said variable bit rate audio encoder (2) comprises:

   means (103) for audio detection and identification (ADI), which in conjunction with audio prediction analysis, predetermine whether an audio frame may be skipped or the appropriate bit rate which should be used for coding of the audio frame, and

   means for computation of the total audio bit rate saving (4) during previous picture coding period and sending the audio bit rate saving (4) to the bit rate allocator (5).

3. A coding system according to claim 2 wherein the said means for audio detection and identification (ADI) comprises:

   means of analysis of frequency and energy content of the input audio signal (1) to determine the energy level and distribution of the energy,

   means of classification of audio signal type from the said energy level and distribution of the energy into silence, predominantly unvoiced or voiced signals, and

   means of estimation of tolerable coding inaccuracy and quantization noise so as to generate a predetermined bit rate for audio coding based on a noise measure or a perceptual measure.

4. A coding system according to any of claims 1 to 3 wherein the values of B_s (4) and B_f (9) are sent to the bit rate allocator (5) when a request signal to the audio encoder (2) and the FIFO buffer (11) is made.

5. A coding system according to any of claims 1 to 4 wherein said variable bit rate video encoder (10) comprises:

   means for dynamically controlling the picture rate according to the available bits for the current picture coding,

   means (31) for controlling the picture coding modes, and

   means for dynamically allocating bits to the current picture according to the picture activity of said picture.

6. A coding system according to claim 5 wherein said dynamic picture coding modes control (31) comprises:

   an intra-picture coding mode which encodes a picture only by the picture contents of its own, the picture coded

BNSDOCID: <EP_____0665693B1_I_>

in this mode is called an I picture, and

decision means for dynamically control the I and P picture coding according to the available bit amount (30) for the current picture and the minimal I picture coding bit rate B_{Imin} and the minimal P picture coding bit rate B_{Pmin}, wherein a picture which is predicted only from the previous picture, is called a P picture.

7. A coding system according to claim 5 or 6 wherein said dynamic picture rate control comprises:

decision means for inserting an extra P picture in between the previous coded picture and the current picture according to the available bit amount (30) for the current picture, the minimal bit amount for a P picture coding and the picture content changes of the picture to be inserted with respect to the previous coded picture.

8. A coding system according to any of claims 1 to 7, wherein said FIFO buffer (11) comprises:

means for calculating the buffer fullness (9) and reporting the value to the bit rate allocator (5) when a request signal from the bit rate allocator (5) is received by the FIFO buffer (11).

9. A coding system according to any of claims 1 to 8, wherein said mulitplexer (14) comprises:

means for retrieving audio and video bitstreams from said FIFO buffer (11) and multiplexing them into a mixed an synchronised bitstream, and

means for generating a constant bit rate multiplexed bitstream output (15) by packetizing the audio and video bitstreams.

Patentansprüche

1. Codiersystem zur Codierung eines Videosignals (6) und eines Audiosignals (1), mit:

einem Audio-Coder (2) mit variabler Bitrate zur Codierung des Audiosignals (1) und zur Bereitstellung einer Menge (4) von eingesparten Bits,

einem Video-Coder (10) mit variabler Bitrate zur Codierung des Videosignals (6),

einem FIFO-Puffer (11) zur Zwischenspeicherung des codierten Audiobitstroms (3) und Videobitstroms (12) und zur Bereitstellung eines Bitstroms (13) mit konstanter Rate und

einer Bitratenzuweisungseinrichtung (5) zur dynamischen Zuweisung einer Menge (8) von Bits für ein aktuelles Bild des zu codierenden Videosignals (6), wobei die Bitratenzuweisungseinrichtung (5) die Menge (8) der Bits für das aktuelle Bild in Abhängigkeit von der Menge (4) der eingesparten Bits des Audio-Coders (2), von der Menge (7) der Bits, die für ein vorhergehendes Bild des Videosignals (6) verwendet wurden, und von einem Füllstand (9) des FIFO-Puffers (11) berechnet,

einem Multiplexer (14) zum Multiplexen des Audio- und Video-Bitstroms (13) mit konstanter Rate mit hinzugefügten Zeitstempeln für eine Synchronisation des Audiosignals und des Videosignals.

2. Codiersystem nach Anspruch 1, wobei der Audio-Coder (2) mit variabler Bitrate umfaßt:

eine Einrichtung (103) zur Audio-Detektion und -Identifikation (ADI), die in Verbindung mit der Audioprädiktionsanalyse entweder festlegt, ob ein Audiorahmen übersprungen werden kann, oder die passende Bitrate festlegt, die zur Codierung des Audiorahmens verwendet werden sollte, und
eine Einrichtung zur Berechnung der gesamten Audiobitrateneinsparung (4) während der vorhergehenden Bildcodierungsdauer und Weitergabe der Audiobitrateneinsparung (4) an die Bitratenzuweisungseinrichtung (5).

3. Codiersystem nach Anspruch 2, wobei die Einrichtung zur Audio-Detektion und -Identifikation (ADI) umfaßt:

eine Einrichtung zur Analyse des Frequenz- und Energiegehaltes des Eingangs-audiosignals (1), um den

BNSDOCID: <EP_____0665693B1_I_>

Energiepegel und die Verteilung der Energie zu bestimmen,

eine Einrichtung zur Klassifikation des Audiosignaltyps in Abhängigkeit von dem Energiepegel und der Verteilung der Energie in Ruhe-, vorwiegend keine-Sprach- oder Sprachsignale und

eine Einrichtung zur Schätzung einer tolerierbaren Codierungenauigkeit und eines tolerierbaren Quantisierungsrauschens, um eine vorherbestimmte Bitrate für die Audiocodierung basierend auf einem Rauschmaß oder einem Wahrnehmungsmaß zu erzeugen.

4. Codiersystem nach einem der Ansprüche 1 bis 3, wobei die Werte von B_s (4) und B_f (9) an die Bitratenzuweisungseinrichtung (5) weitergegeben werden, wenn ein Anforderungssignal an den Audio-Coder (2) und den FIFO-Puffer (11) abgegeben ist.

5. Codiersystem nach einem der Ansprüche 1 bis 4, wobei der Video-Coder (10) mit variabler Bitrate umfaßt:

eine Einrichtung zur dynamischen Steuerung der Bildrate in Abhängigkeit von den verfügbaren Bits für die aktuelle Bildcodierung,

eine Einrichtung (31) zur Steuerung der Bildcodierbetriebsarten, und

eine Einrichtung zur dynamischen Zuweisung von Bits für das aktuelle Bild in Abhängigkeit von der Bildaktivität des Bildes.

6. Codiersystem nach Anspruch 5, wobei die dynamische Steuerung (31) der Bild codierbetriebsarten umfaßt:

eine Intra-Bildcodierbetriebsart, die ein Bild nur basierend auf dessen Bildinhalt codiert, wobei das in dieser Betriebsart codierte Bild ein I-Bild genannt wird, und

eine Entscheidungseinrichtung zur dynamischen Steuerung der I- und P-Bildcodierung in Abhängigkeit von der verfügbaren Bitmenge (30) für das aktuelle Bild und die minimale I-Bildcodierbitrate B_{Imin} und die minimale P-Bildcodierbitrate B_{Pmin}, wobei ein Bild, wenn es nur von dem vorhergehenden Bild prädiziert wird, ein P-Bild genannt wird.

7. Codiersystem nach Anspruch 5 oder 6, wobei die dynamische Bildratensteuerung umfaßt:

eine Entscheidungseinrichtung zur Einfügung eines zusätzlichen P-Bildes zwischen das vorhergehend codierte Bild und das aktuelle Bild in Abhängigkeit von der verfügbaren Bitmenge (30) für das aktuelle Bild, der minimalen Bitmenge für eine P-Bildcodierung und den Bildinhaltsänderungen des Bildes, das bezüglich zu dem

vorhergehend codierten Bild einzufügen ist.

8. Codiersystem nach einem der Ansprüche 1 bis 7, wobei der FIFO-Puffer (11) umfaßt:

eine Einrichtung zur Berechnung des Puffer-Füllstandes (9) und zur Weitergabe des Wertes an die Bitratenzuweisungseinrichtung (5), wenn ein Anforderungssignal von der Bitratenzuweisungseinrichtung (5) von dem FIFO-Puffer (11) empfangen wird.

9. Codiersystem nach einem der Ansprüche 1 bis 8, wobei der Multiplexer (14) umfaßt:

eine Einrichtung zum Abrufen der Audio- und Videobitströme von dem FIFO-Puffer (11) und zum Multiplexen von ihnen in einen gemischten und sychronisierten Bitstrom, und

eine Einrichtung zur Erzeugung einer gemultiplexten Bitstromausgabe (15) mit einer konstanten Bitrate durch Packetisierung der Audio- und Videobitströme.

BNSDOCID: <EP_____0665693B1_I_>

Revendications

1. Système de codage pour coder un signal vidéo (6) et un signal audio (1), comportant :

   un codeur audio à débit binaire variable (2) pour coder ledit signal audio (1) et délivrer une quantité (4) de bits sauvegardés,
   un codeur vidéo à débit binaire variable (10) pour coder ledit signal vidéo (6),
   un tampon FIFO (11) pour mémoriser le flot de bits audio (3) et le flot de bits vidéo (12) codés et délivrer un flot de bits de débit constant (13), et
   une unité d'assignation de débit binaire (5) pour assigner d'une manière dynamique une quantité (8) de bits à une image courante dudit signal vidéo (6) à coder, ladite unité d'assignation de débit binaire (5) calculant ladite quantité (8) de bits à utiliser pour l'image courante en fonction de la quantité (4) desdits bits sauvegardés du codeur audio (2), de la quantité (7) de bits utilisés pour une image précédente dudit signal vidéo (6) et d'un taux de remplissage (9) dudit tampon FIFO (11),
   un multiplexeur (14) pour multiplexer le flot de bits audio et vidéo de débit constant (13) avec des estampilles temporelles ajoutées pour une synchronisation du signal audio et du signal vidéo.

2. Système de codage selon la revendication 1, dans lequel ledit codeur audio à débit binaire variable (2) comporte :

   des moyens (103) de détection et d'identification audio (ADI) qui, en association avec une analyse prédictive audio, prédéterminent si une trame audio peut être sautée ou le débit binaire approprié à utiliser pour coder la trame audio, et
   des moyens pour calculer l'économie de débit binaire audio totale (4) durant la période de codage d'image précédente et envoyer l'économie de débit binaire audio (4) à l'unité d'assignation de débit binaire (5).

3. Système de codage selon la revendication 2, dans lequel lesdits moyens de détection et d'identification audio (ADI) comportent :

   des moyens pour analyser le contenu fréquentiel et énergétique du signal audio d'entrée (1) afin de déterminer le niveau d'énergie et la distribution de l'énergie,
   des moyens de classification du type de signal audio à partir dudit niveau d'énergie et de ladite distribution de l'énergie en signal de silence, signal majoritairement sans voix ou signal de voix, et
   des moyens pour estimer l'imprécision de codage et le bruit de quantification admissibles de manière à générer un débit binaire prédéterminé pour le codage audio sur la base d'une mesure du bruit ou d'une mesure de sensibilité.

4. Système de codage selon l'une quelconque des revendications 1 à 3, dans lequel les valeurs de B_s (4) et B_f (9) sont envoyées à l'unité d'assignation de débit binaire (5) lorsqu'un signal de demande est envoyé au codeur audio (2) et au tampon FIFO (11).

5. Système de codage selon l'une quelconque des revendications 1 à 4, dans lequel ledit codeur vidéo à débit binaire variable (10) comporte :

   des moyens pour commander d'une manière dynamique le débit d'image en fonction des bits disponibles pour le codage de l'image courante,
   des moyens (31) pour commander les modes de codage d'image, et
   des moyens pour assigner d'une manière dynamique des bits à l'image courante en fonction de l'activité d'image de ladite image.

6. Système de codage selon la revendication 5, dans lequel ladite commande dynamique des modes de codage d'image (31) comporte :

   un mode de codage intra-image qui code une image uniquement en fonction de son contenu, l'image codée dans ce mode étant appelée image I, et
   des moyens de décision pour commander d'une manière dynamique le codage d'une image I et P en fonction de la quantité de bits disponible (30) pour l'image courante et du débit binaire minimal de codage d'une image I B_{Imin} et du débit binaire minimal de codage d'une image P B_{Pmin}, une image prédite uniquement à partir de l'image précédente étant appelée image P.

BNSDOCID: <EP_____0665693B1_I_>

7. Système de codage selon la revendication 5 ou 6, dans lequel ladite commande dynamique de débit d'image comporte :

des moyens de décision pour insérer une image P supplémentaire entre l'image codée précédente et l'image courante en fonction de la quantité de bits disponible (30) pour l'image courante, de la quantité minimale de bits nécessaire à un codage d'image P et des variations du contenu d'image de l'image à insérer par rapport à l'image codée précédente.

8. Système de codage selon l'une quelconque des revendications 1 à 7, dans lequel ledit tampon FIFO (11) comporte :

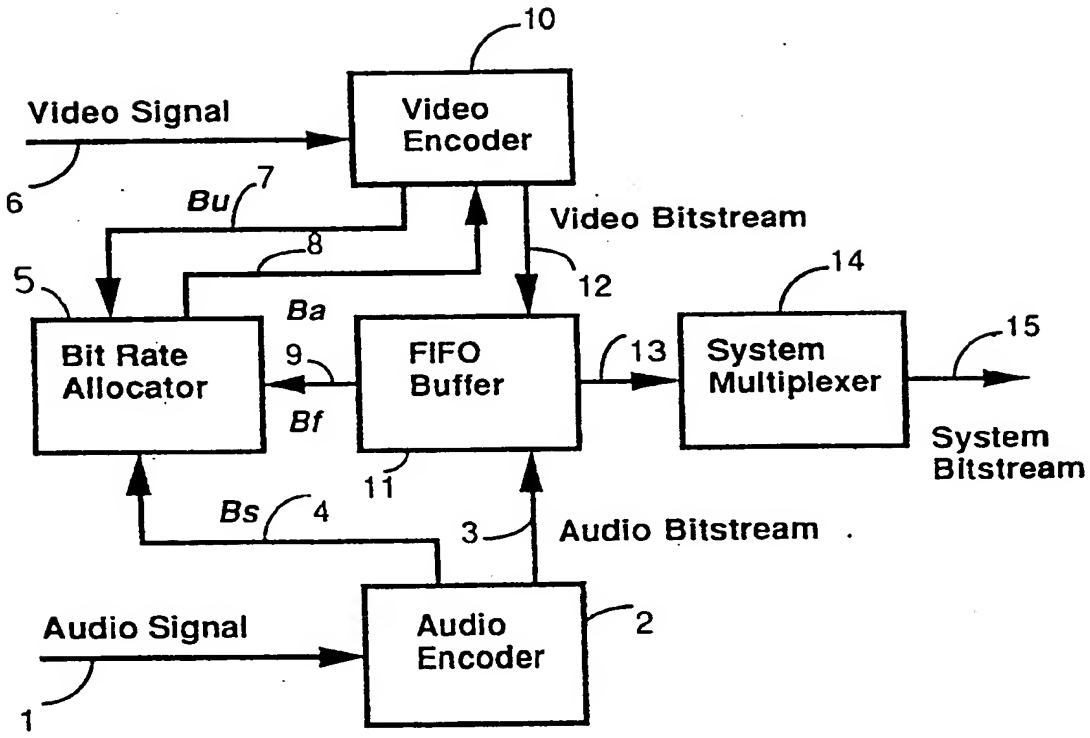
des moyens pour calculer le taux de remplissage du tampon (9) et pour transmettre la valeur à l'unité d'assignation de débit binaire (5) lorsqu'un signal de demande en provenance de l'unité d'assignation de débit binaire (5) est reçu par le tampon FIFO (11).

9. Système de codage selon l'une quelconque des revendications 1 à 8, dans lequel ledit multiplexeur (14) comporte :

des moyens pour extraire les flux de bits audio et vidéo à partir dudit tampon FIFO (11) et les multiplexer sous la forme d'un flux de bits synchronisés mélangés, et
des moyens pour générer une sortie à flux de bits multiplexés à débit binaire constant (15) en paquetisant les flux de bits audio et vidéo.

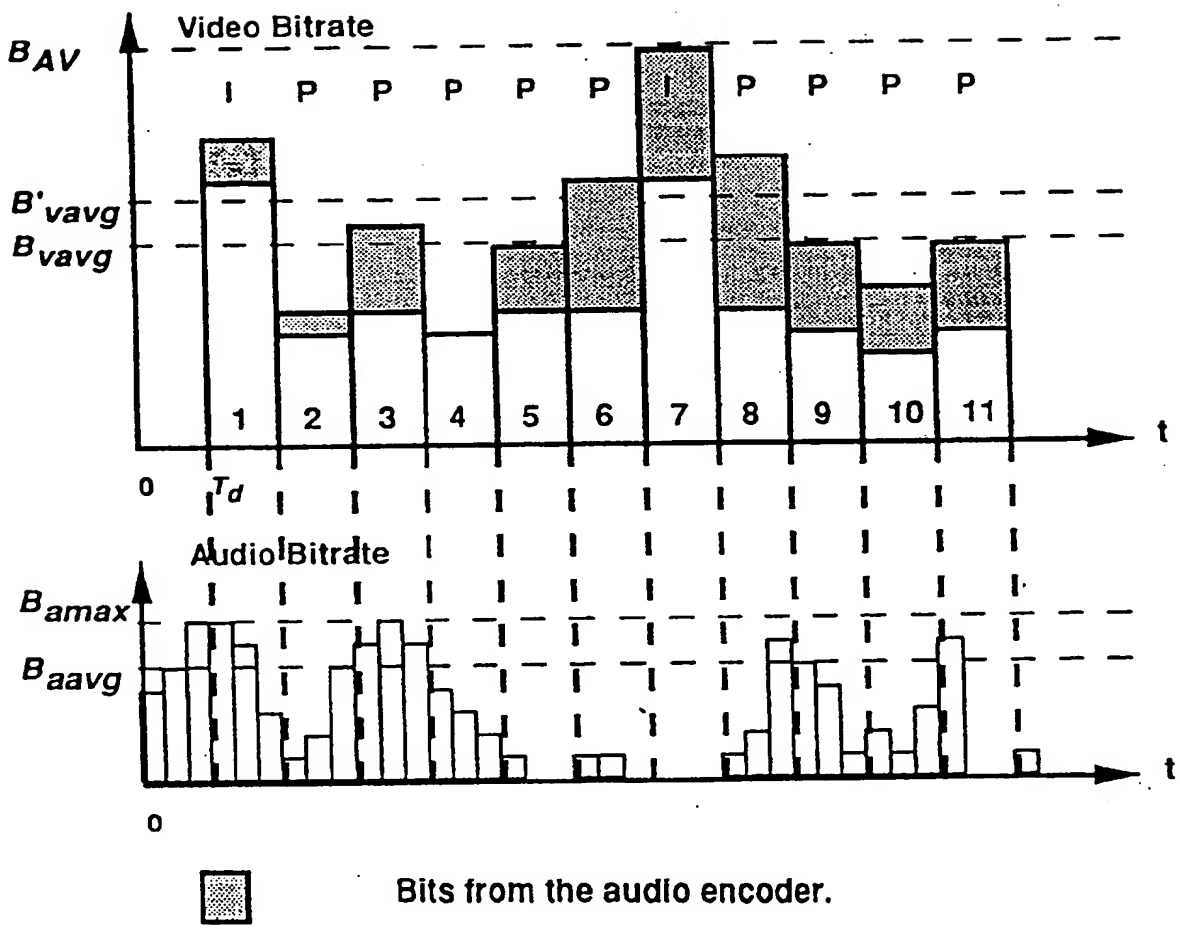
BNSDOCID: <EP_____0665693B1_I_>

Fig. 1



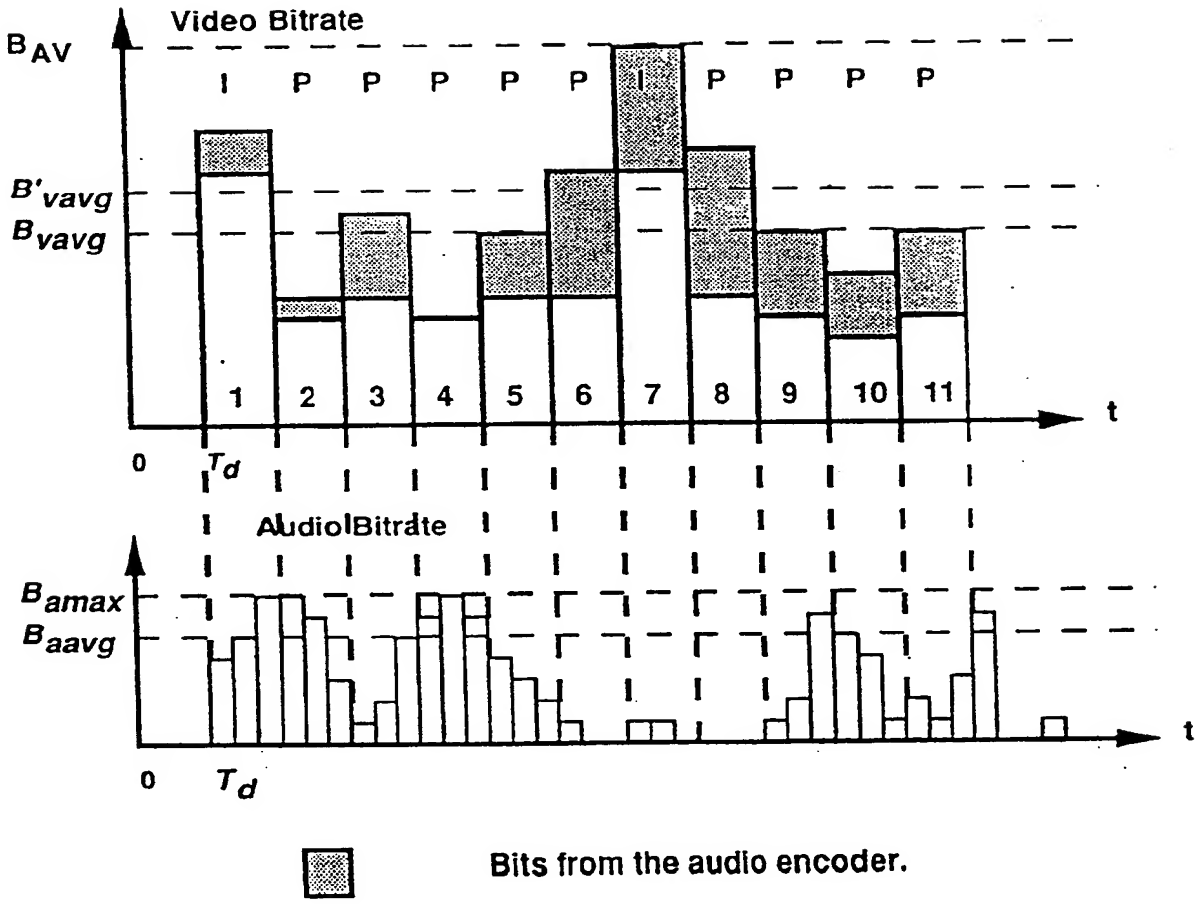
BNSDOCID: <EP_____0665693B1_I_>

Fig. 2



BNSDOCID: <EP_____0665693B1_I_>

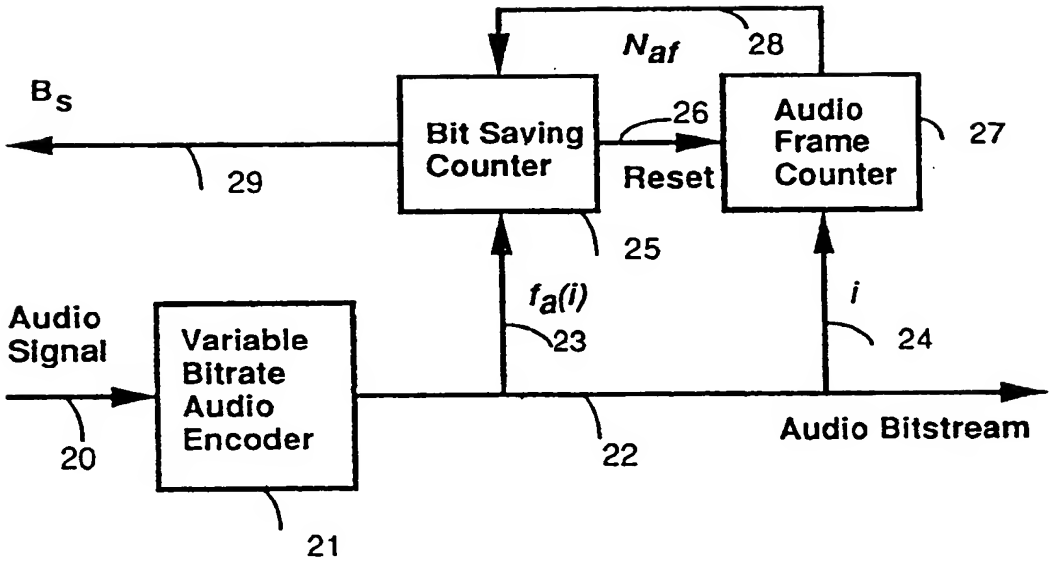
Fig. 3



Bits from the audio encoder.

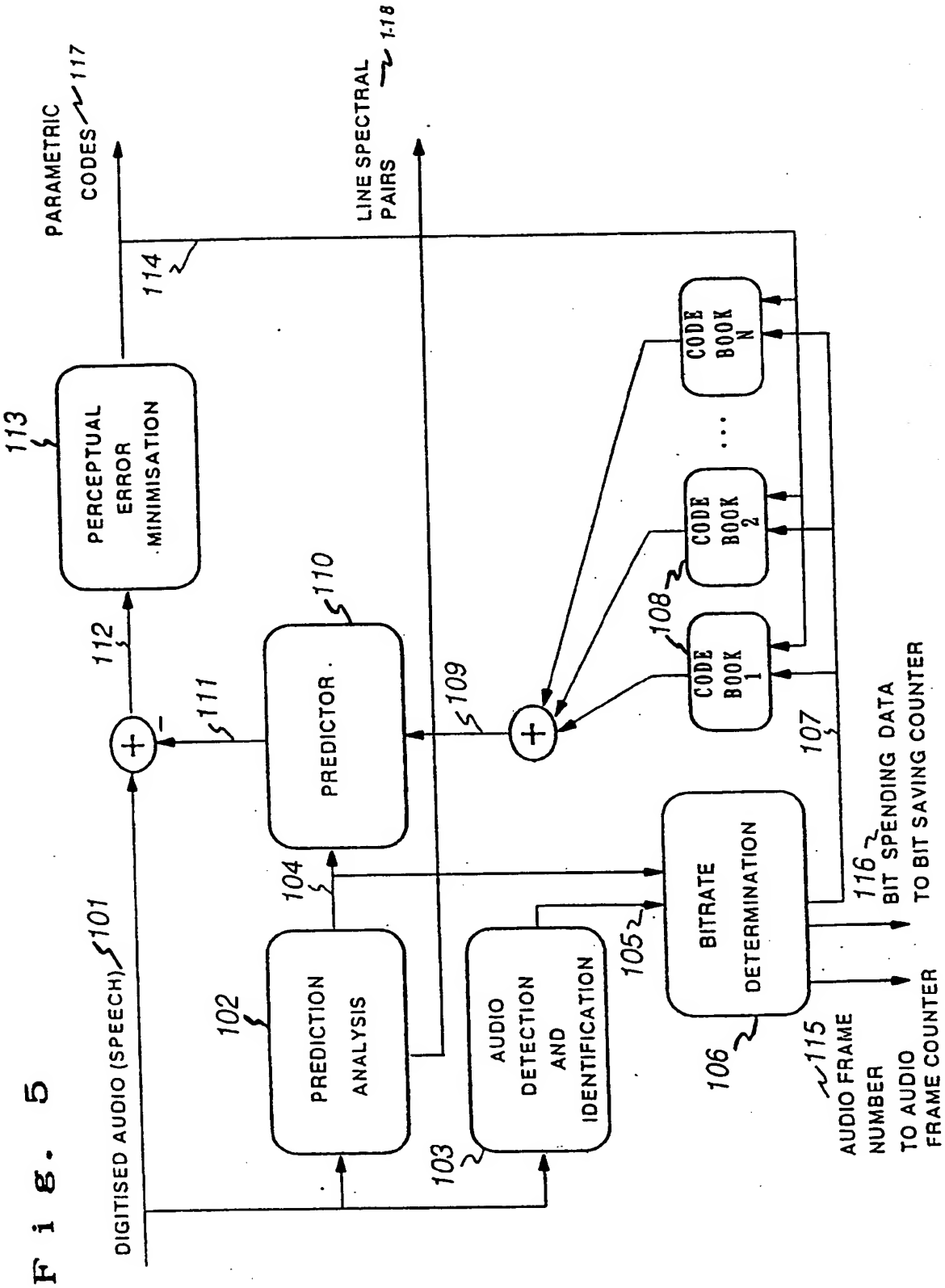
BNSDOCID: <EP_____0665693B1_I_>

Fig. 4



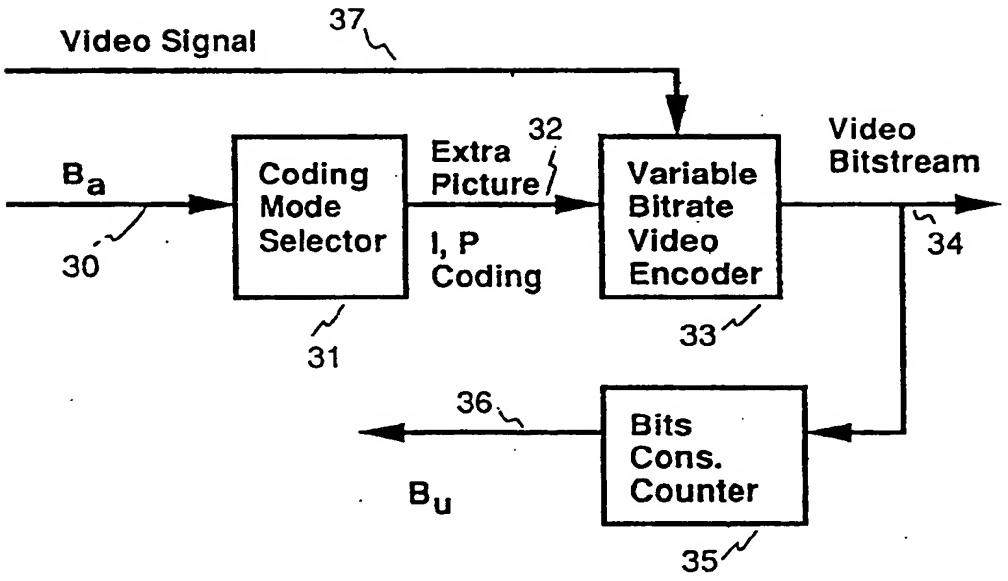
BNSDOCID: <EP_____0665693B1_I_>

Fig. 5



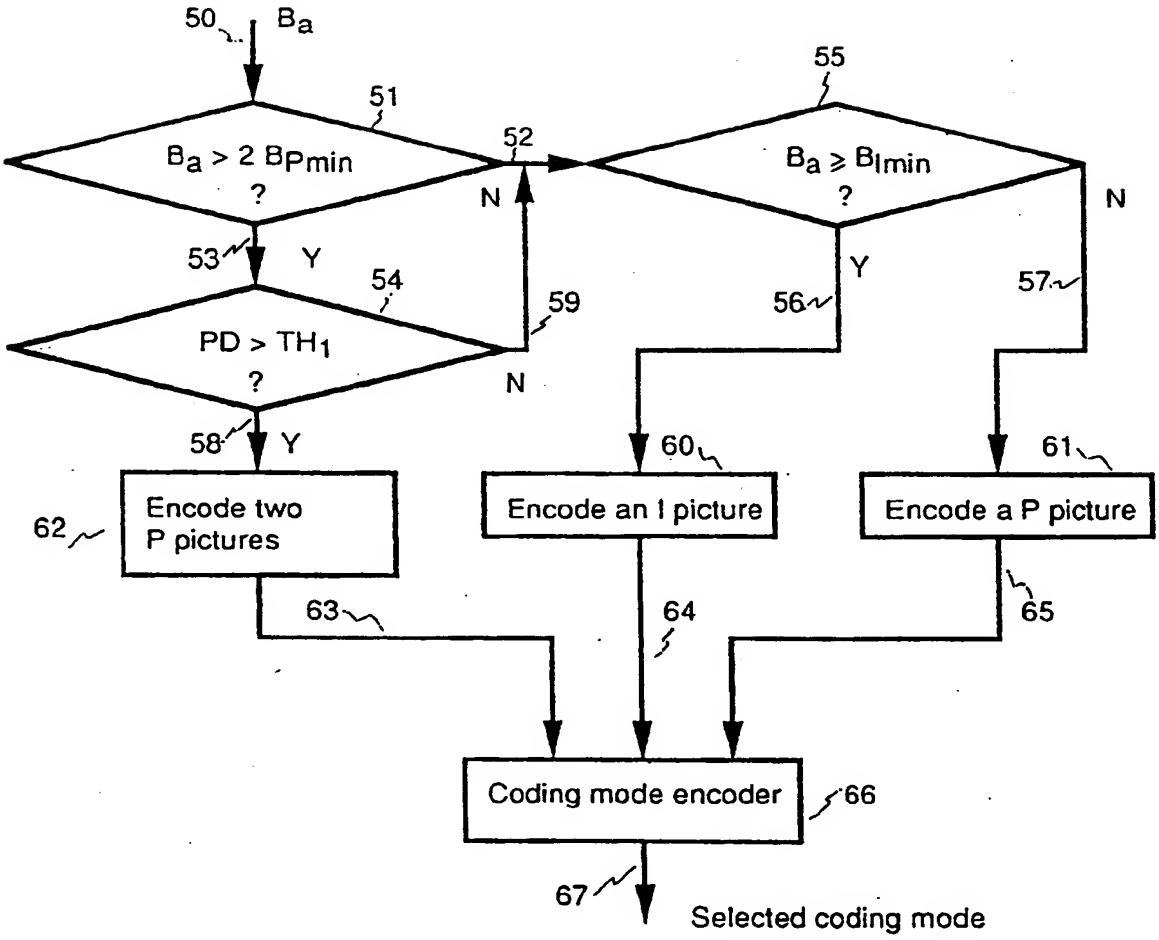
BNSDOCID: <EP_____0665693B1_I_>

Fig. 6



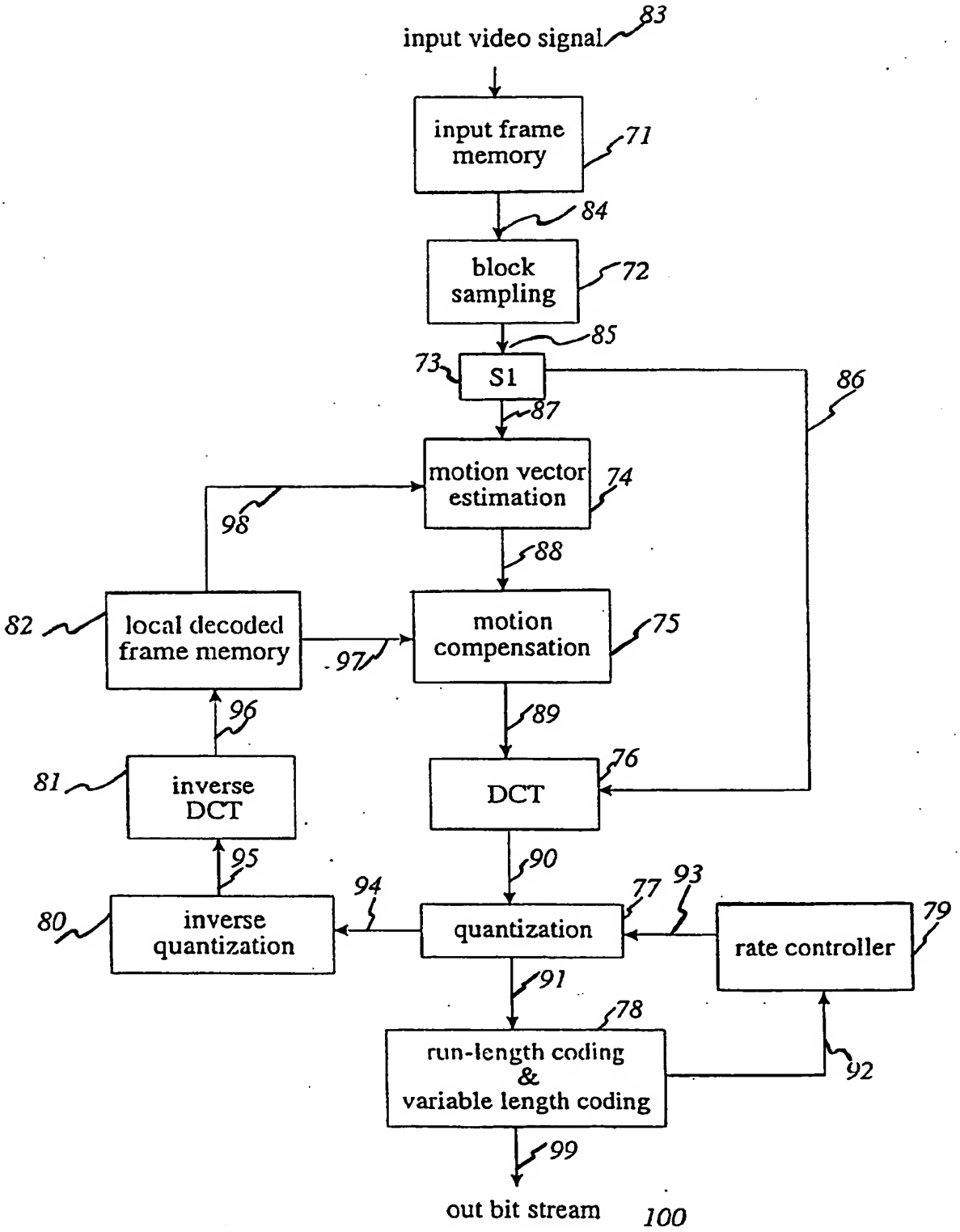
BNSDOCID: <EP_____0665693B1_I_>

Fig. 7



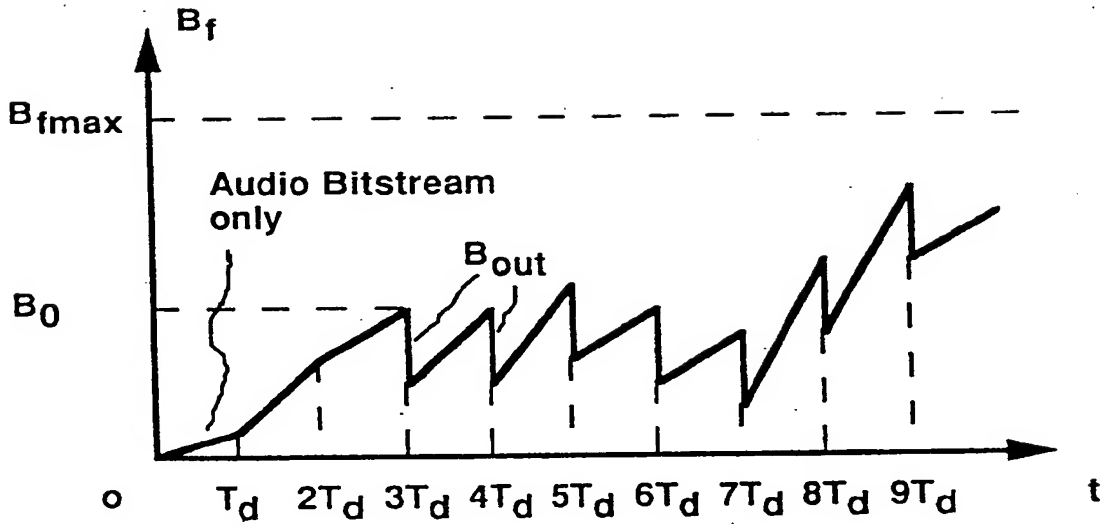
BNSDOCID: <EP_____0665693B1_I_>

Fig. 8



BNSDOCID: <EP_____0665693B1_I_>

Fig. 9



BNSDOCID: <EP_____0665693B1_I_>